

A Visual Analytics Approach to Debugging Cooperative, Autonomous Multi-Robot Systems' Worldviews

Suyun "Sandra" Bae*

Federico Rossi†

Joshua Vander Hook‡

Scott Davidoff§

Kwan-Liu Ma¶

*¶ University of California, Davis

†‡§ Jet Propulsion Laboratory, California Institute of Technology

ABSTRACT

Autonomous multi-robot systems, where a team of robots shares information to perform tasks that are beyond an individual robot's abilities, hold great promise for a number of applications, such as planetary exploration missions. Each robot in a multi-robot system that uses the shared-world coordination paradigm autonomously schedules which robot should perform a given task, and when, using its *worldview*—the robot's internal representation of its belief about both its own state, and *other* robots' states. A key problem for operators is that robots' worldviews can fall out of sync (often due to weak communication links), leading to desynchronization of the robots' scheduling decisions and inconsistent emergent behavior (e.g., tasks not performed, or performed by multiple robots). Operators face the time-consuming and difficult task of making sense of the robots' scheduling decisions, detecting de-synchronizations, and pinpointing the cause by comparing every robot's worldview. To address these challenges, we introduce MOSAIC Viewer, a visual analytics system that helps operators (i) make sense of the robots' schedules and (ii) detect and conduct a root cause analysis of the robots' desynchronized worldviews. Over a year-long partnership with roboticists at the NASA Jet Propulsion Laboratory, we conduct a formative study to identify the necessary system design requirements and a qualitative evaluation with 12 roboticists. We find that MOSAIC Viewer is faster- and easier-to-use than the users' current approaches, and it allows them to stitch low-level details to formulate a high-level understanding of the robots' schedules and detect and pinpoint the cause of the desynchronized worldviews.

Keywords: Multi-Robot Systems, Human-Subjects Qualitative Studies, Debugging.

Index Terms: I.3.8 [Computer Graphics]: Applications—

1 INTRODUCTION

Autonomous multi-robot systems (MRS) are systems with two or more autonomous robots (often referred to as *agents*), that coordinate and share information so as to perform tasks cooperatively. This cooperation, in particular, drives interest in their potential to perform highly complex tasks in diverse contexts from search and rescue in hazardous environments [4, 23, 47] to team sports [34, 51], and even space missions [56, 62]. The complexity of these systems also introduces a problem of usability for operations—or operability. Researchers must monitor the behavior of individual agents as well as behavior that emerges from their cooperation [32] and see how small changes to their systems affect the overall system performance.

One area where this emergent complexity can be particular challenging for operators is *distributed scheduling* [52], i.e., cooperatively deciding which agent should perform a given task and when. To track even a single task on a single agent, operators need to understand and track inter-task dependencies the precedence constraints that the agents' schedule must satisfy (e.g., scientific data must be collected before it is analyzed and transmitted to Earth). With distributed scheduling, the effort required to understand the state of the system increases quadratically, as tasks assigned to one agent might be shared with or even performed by others. In addition, the overall system's behavior depends not only on each agent's individual state but also on each agent's belief about the state of other agents and of the environment (i.e., a *worldview* [30]).

Agents' worldviews introduce a second challenge that further complicates tracking agents' tasks, as worldviews can fall out of sync (often due to weak communication links). This leads to desynchronization of the agents' scheduling decisions and inconsistent emergent behavior (e.g., tasks not performed, or performed by multiple agents). To debug these inconsistent behaviors, operators must pinpoint the source of the desynchronization by comparing every agent's worldview. This process is not only critical for debugging and failure detection purposes, but also enormously time-consuming and difficult: operators need to examine the high-dimensional product of every attribute in every agent's worldview.

While previous research has explored ways to represent the views of single agents using text [14, 22, 49] or superimposed over videos [2, 69], we explore how a visual analytics approach [11] can encode the belief agents have of themselves and about the state of other agents. To that end, we engaged in a year-long collaboration with a team of MRS researchers and operators at the NASA Jet Propulsion Laboratory (NASA JPL). The collaboration began with a 10-week formative investigation to identify the core challenges of distributed scheduling, utilizing the MOSAIC distributed scheduling framework [75] as a laboratory to explore this objective. This work was followed by six months of iterative co-design with a core MOSAIC team member to produce MOSAIC Viewer, a visual analytics application that helps operators (i) make sense of the agents' schedules and (ii) detect and conduct a root cause analysis of the desynchronized worldviews. To compare worldviews, MOSAIC Viewer draws inspiration from the *diff* algorithm, which is commonly used for text comparison [35] to emphasize the differences of agents' worldview. Lastly, we demonstrate the effectiveness of our method and system with two case studies and evaluate the application through a qualitative study with 12 roboticists at JPL. The study reveals MOSAIC Viewer is easier- and faster-to-use than the users' current text-based approaches. The study helps to explain how applications like MOSAIC Viewer can support worldview desynchronization debugging. In particular, from our evaluation, we find that our tool supports the following practices:

- System speed and interactivity streamline higher-level analyses;
- Trust for summary displays grew with experience;
- Knowing how is not enough—users need to know “why” in

*e-mail: suybae@ucdavis.edu

†e-mail: federico.rossi@jpl.nasa.gov

‡e-mail: hook@jpl.nasa.gov

§e-mail: scott.davidoff@jpl.nasa.gov

¶e-mail: klma@ucdavis.edu

Table 1: Attributes in every agent’s worldview. For every attribute, an agent has a value for itself and the presumed values for the other agents.

Attribute	Worldview Attribute Value		Data Structure
	Self	Others	
Location	The robot’s location	Presumed location of other robots	2D Coordinates
Science Zone	The robot’s classification of whether it is in a science zone	Presumed classification of whether other robots are in science zones	Boolean Array
Battery Level	The robot’s battery	Presumed battery level of other robots	Ordinal Array
CPU Utilization	The robot’s CPU level	Presumed CPU level of other robots	Ordinal Array
Actions	The actions the robot is currently performing	Actions other robots are believed to be performing	Event Sequence
Communication	Bandwidth between self and other robots	Presumed bandwidths between other pairs of robots	Graph

order to back trace the root causes of the problem;

- Different sets of assumptions affect data interpretation.

This particular design study [63] helps researchers understand the fit between the problem of MRS operators debugging desynchronized worldviews and MOSAIC Viewer. In this paper we contribute: (i) a set of system design requirements based on a year-long formative study with domain experts in multi-robot systems; (ii) a visual analytics tool that helps operators understand and compare agents’ worldviews with a comparison technique inspired from the *diff* algorithm; and (iii) we characterize how the system supports effective troubleshooting, with evidence gathered from a study of the system.

2 BACKGROUND

To motivate and situate our work, we first discuss the specific challenges of supervising autonomous MRS and unpack the details of the MOSAIC distributed scheduling framework [75] that we use to explore desynchronized worldview debugging.

2.1 Autonomous Multi-Robot Systems

In contrast to multi-agent systems [46] which are enacted entirely as software, in this work we focus on multi-robot systems that have to negotiate with real-world constraints (e.g., limited and time-varying communication bandwidth and dynamic battery levels) that are often not considered in the multi-agent systems literature [10, 38, 44].

While some MRS researchers investigate MRS that utilize explicit [65] or centralized [78] coordination, we focus on MRS that use *shared-world* coordination [3, 18], which have proven to be highly popular in field applications [7, 58] due to its simplicity, scalability, and resiliency. In this approach, every robotic agent has a *worldview* [30]—an internal representation of the world and of the other agents’ states—that is updated through constant communication with other agents. In our case, Table 1 summarizes the different attributes found in an agent’s worldview within the MOSAIC framework (described in Sect. 2.2). Based on its own worldview, each agent independently computes the optimal strategy for *all* agents, and then executes its own part of the computed strategy. If all agents have the same worldview, this results in coordinated behavior.

While a shared-world approach introduces many benefits, a key complexity it introduces is that, if the agents are unable to communicate with each other, their worldviews can fall out of sync, resulting in uncoordinated decisions (e.g., a task may be performed by two agents, or it may not be performed at all). This issue is especially problematic in harsh environments such as underground caves [17, 42], where ensuring constant reliable communication is infeasible. Therefore, in order to understand the overall behavior of MRS, it is critical to understand the worldview of each agent. Furthermore, in order to mitigate the effect of worldview desynchronization, an operator must be able to identify the cause (e.g., slow propagation of information on low-bandwidth data links or the failure of an agent’s radio) to plan for corrective action. The concurrent, distributed, and complex components of MRS makes the debugging process significantly difficult, and previous research has identified that these tasks require considerable attention [25] and would benefit from more appropriate, effective tools [73].

2.2 MOSAIC Distributed Scheduling

Within the MOSAIC distributed scheduling framework, each agent can perform a set of **navigation** tasks and **science** tasks. Navigation tasks, which model activities such as localization and path planning, are *mandatory*, and all agents must perform them. Science tasks, which model collection and analysis of scientific observables, are *optional*. Though individual science tasks are not required for mission success, the objective of the MOSAIC framework is to perform as many science tasks as possible. However, an agent can perform science tasks only if it has the time and energy resources to also guarantee the execution of the mandatory navigation tasks.

Both navigation tasks and science tasks have precedence constraints, enforcing that tasks must be accomplished in sequence. With science tasks, performing analysis of scientific measurements requires that the data be collected first. With navigation tasks, performing localization through visual odometry requires collecting camera images first. Hence, we will refer to navigation tasks and science tasks as a “chain of tasks” (i.e., several tasks with a chain of inter-task dependencies). A key advantage of MRS is that agents do not need to accomplish each task all by themselves—they may receive assistance from other agents. An agent in a science zone may request assistance with its navigation tasks in order to free up computational resources for science tasks. Certain computational tasks, such as performing visual odometry and analyzing data, are relocatable to other agents. However, not all tasks are relocatable—for instance, tasks requiring the use of an agent’s hardware resources (e.g., capturing images or collecting scientific measurements) are not. The MOSAIC scheduler takes all these constraints into account and computes (i) what optional tasks can be performed and (ii) which agents should perform relocatable tasks based on the agents’ capabilities and communication links between the agents [75].

In this paper, we consider datasets with ten agents generated by running the MOSAIC scheduler in the loop with a multi-robot simulator that captures the availability of science zones, robot battery levels, and communication link bandwidths, which reflects the standard practice in robotics research [53, 74, 75]. This level of simulation fidelity is well-matched with the level of abstraction at which MOSAIC operates; while field testing may present different underlying causes for worldview desynchronization, the effect on the data used in this paper (i.e., disagreement between the agents’ world views) would be indistinguishable from the output of the simulations. Accordingly, the use of a simulator has a negligible impact on the fidelity of worldview debugging. In these datasets, each agent has six attributes in its worldview (Table 1) and is endowed with an agent ID. Each agent wishes to perform three mandatory navigation tasks; agents in “science zones” also wish to perform three optional science tasks. Each set of tasks has a chain of dependency constraints. One agent—the base station—is a special agent that does not need to perform the navigation or science tasks and is equipped with a faster processor. Its purpose is to help the other agents with its computing capabilities. Lastly, we remark that the scale of ten agents is representative of proposed extra-planetary (i.e., outside of Earth) MRS mission concepts under consideration for the next decade [8, 39, 40, 56].

Table 2: Summary of the participants’ background, current tools and the extent they participated in the year-long formative study.

Participants	MOSAIC Affiliation	Years of Full-Time Professional Experience	Current Tools					Participation			
			PLT ¹	RT ²	CLP ³	DBT ⁴	DB-LT ⁵	FS ⁶	Co-Design	FE ⁷	User Study
P0	Core	1 – 5 years	✓	✓	✓	–	–	✓	✓	✓	–
P1	Core	1 – 5 years	–	✓	–	–	–	✓	–	–	✓
P2	Core	Less than 1 year	–	–	–	–	✓	–	–	–	✓
P3	Non-Core	5 – 10 years	✓	–	–	–	–	✓	–	–	✓
P4	Non-Core	Less than 1 year	✓	–	✓	–	–	–	–	–	✓
P5	Core	10 – 15 years	–	–	✓	–	–	✓	–	✓	✓
P6	Non-Core	1 – 5 years	✓	✓	–	–	–	✓	–	–	✓
P7	Non-Core	1 – 5 years	–	✓	✓	✓	–	–	–	–	✓
P8	Core	1 – 5 years	✓	–	✓	–	–	✓	–	–	✓
P9	Non-Core	1 – 5 years	–	✓	–	–	–	–	–	–	✓
P10	Non-Core	More than 15 years	–	✓	✓	–	–	–	–	–	✓
P11	Non-Core	1 – 5 years	–	✓	–	–	–	–	–	–	✓
P12	Core	10 – 15 years	–	–	✓	–	–	✓	–	✓	✓

¹ MATLAB/Matplotlib (ad-hoc scripting languages and plotting tools); ² RVIZ and ROS-based plotting tools [55]; ³ Command-line (CLI) logging tools;⁴ Command-line debugging and backtracing tools (GDB); ⁵ Database-backed logging tools; ⁶ Formative Study; ⁷ Formative Evaluation

3 RELATED WORK

We present MRS supervision tools used in the robotics community and promising visualization techniques for worldview debugging.

3.1 Scheduling and Timeline Views in Robotics

In addition to the various multi-purpose robotics toolkits [55], specialized visual analytics tools have been developed to track robots’ task completion [49, 68, 70]. These tools include timeline views that are often organized around two underlying data types. First, we observe “agent-centric” (AC) timelines [12, 36], which map timeline rows to individual agents, foregrounding the tasks performed by each agent (either for themselves or to assist other agents). In contrast, the second approach, “task-centric” (TC) [59], organizes timeline rows around tasks and their dependencies and focuses on when tasks are completed, rather than *who* is executing them.

We find AC timelines to be an incomplete solution for the problem of distributed scheduling, where tasks offloaded to other agents can be difficult to trace. We find a similar limitation with TC timelines for distributed scheduling problems, where backtracing tasks with dependencies can be difficult to explain failures. To that end, MOSAIC Viewer uses an AC-TC hybrid timeline. The timeline includes glyphs that visually encapsulate the completion status of individual tasks, and foregrounds task dependencies using interactions [21, 37, 77] (see Sect. 5.3 for more details about the timeline). Even with these adaptations, the timeline view is necessary but not sufficient to support the many tasks robotics researchers and operators face when debugging unexpected behaviors due to worldview desynchronization. To complete the application, we turn our attention to views designed to support worldview debugging.

3.2 Worldview Debugging

To mitigate the effect of worldview desynchronization, an operator must be able to identify the root cause of the desynchronization condition (Sect. 2.1). Researchers have investigated methods that display worldview state variables using structured text, through logfile analysis [22] and watchpoints [14], similar to those found in software development IDE’s. These initial explorations provide utility, but at the same time require considerable attention and focus [54, 72]. One objective of our research is to explore a more expressive and lower cognitive load approach for users to examine the high-dimensional product of agents’ worldviews. Visual tracking [2, 69], which overlays line graphs [2] and glyphs [69] that describe the agent’s state on top of a video of agents performing tasks, is another approach. While this approach can effectively show the state of individual agents, MRS worldview debugging requires operators to understand agents’ beliefs about *other* agents’ states as well.

To the best of our knowledge, we know of no research within the visualization community that has explicitly looked at the problem domain of distributed MRS worldview debugging. However, researchers have explored various representations and interaction

strategies for data with similar underlying representations. For example, as worldviews have multiple attributes, we build on visual comparison techniques rooted in multivariate data research [43]. In particular, we rely on visual comparison techniques [26] that include explicit encoding to represent the system’s consensus of an agent’s view of its own state, and juxtaposition to highlight the differences among worldviews. The same strategies to compare parameters of a dataset have been applied across a number of domains, from malware sampling [27] to time series data [50].

Another thread that defines analytics tools that perform multivariate comparisons is the actual algorithm they select to highlight parameter differences. DiffMatrix [66], for example, highlights the difference between two parameter values using the arithmetic subtraction operators, while OnSet [60] uses the union and intersection set operators. Our work, like Vdiff [6] and TACO [50], applies the diff algorithm [35] from text processing (see Sect. 5.2 for details). The system we present, overall, extends the work on multivariate comparison into the domain of distributed MRS worldview debugging and contributes a detailed analysis of the fit of our approach to that domain. We now turn our attention to the formative research that informed our understanding of the problem.

4 FORMATIVE STUDY

MOSAIC Viewer is the product of a year-long engagement, organized around three distinct phases, with its users. Table 2 summarizes the phases of the engagement and describes how each of the 13 participants engaged with this project. (Note: P0 is a superuser who guided the design process, but did not participate in the user study).

This section describes the first and second phase: a 10-week formative study with 7 domain experts and a 6-month co-design study with 1 domain expert. The third phase (Sect. 6.1) is a formal user study that evaluates how well MOSAIC Viewer supports visual debugging for worldview desynchronization.

4.1 Approach

The objectives of the first phase, a formative study, was to gain a deeper understanding of the core challenges of distributed scheduling and the users’ needs. The study was organized around an initial contextual inquiry [33], which allowed the research team to observe the roboticists at work in their own environments. For four weeks, the protocol consisted of six sessions of 90 minutes of semi-structured interviews and an artifact walk-through where roboticists shared the tools and processes that define their MRS work practice. The research team took notes, and captured images and video recordings to highlight observations for post-analysis.

With a preliminary understanding of the problem domain, the research team then elaborated 12 paper prototypes [57] in order to test initial ideas early and quickly. After four rounds of user testing with all users, these prototypes evolved to higher-fidelity code-based prototypes with real MRS data. From these prototyping sessions, we

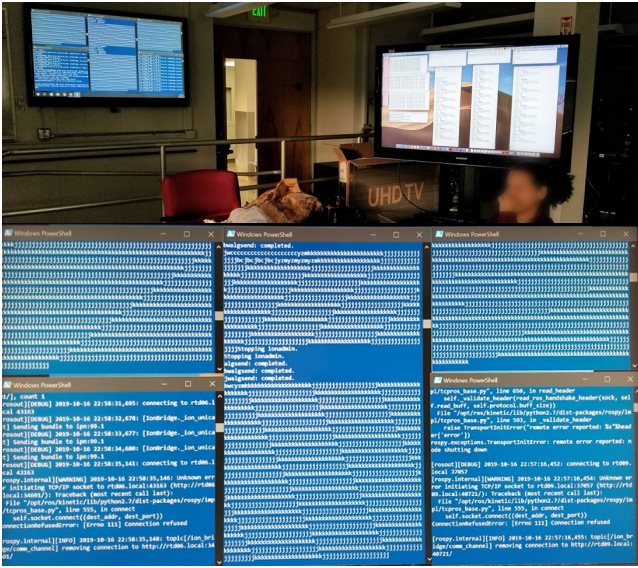


Figure 1: A 50-inch and a 70-inch display. The 15 terminal windows allows P1 and 2 other researchers (not pictured) to debug 3 robots.

identified that comparing agents’ worldviews is a core component we needed to address in order to help roboticists make sense of the agents’ schedules. Then in the second phase, we engaged in 6 months of detailed co-design sessions [29] with P0 to unpack the specific challenges of worldview comparisons. We found that no existing visualizations met the needs of the problem, and we collaborated with P0 to iteratively prototype different visual representations that compare agents’ worldviews. The past designs can be found in the supplementary material. After identifying a potential design, we conducted a formative evaluation [16] with three roboticists, where we revised the system design based on their feedback, such as introducing the hierarchical interactions that we list in Sect. 5.3.

4.2 Findings

The materials captured during phase 1 were annotated with themes, which were then grouped using an emergent theme analysis [1].

One key analysis takeaway reveals roboticists invested substantial amounts of effort and time to track the state of each agent’s worldview in order to interpret which tasks are scheduled for which robot. To illustrate, we observed examples where roboticists would need to open 3 or 4 terminal windows of messaging logs per agent and then would recruit multiple expert colleagues who would each be tasked to monitor a single agent on their own screens (Fig. 1). To elaborate a shared understanding of all the worldviews, including the divergences, the roboticists would collectively, as P1 shares, “shout out the state of what they’re seeing” as they visually scan their respective displays to find relevant information.

Roboticists must first assess worldview synchronization as this dictates the flow for the rest of their analysis. If roboticists conclude that the worldviews are synchronized, they would then assess the system’s performance (i.e., how many tasks are being performed, and by whom). To perform this objective, we observed researchers first engaged in a TC perspective to quickly determine the state of the agent’s task: they did not seek the low-level details of how a task has been accomplished. Conversely, if an agent failed, researchers switched to an AC perspective to seek for the low-level details of an agent’s activities that explained the failure. However, if roboticists concluded that agents’ worldviews are desynchronized, they would transition into a debugging mode to plan for corrective action (i.e., determine who is out of sync and reason why the desynchronization occurred). Roboticists would read out single states within each of their robots’ worldviews again and check for agreement about each state before moving onto the next worldview.

Table 3: Researcher’s goals and subgoals.

Goal	Subgoal
Assess worldview synchronization	1. If desynchronized, who is out of sync with whom? 2. What is the cause of the desynchronization?
Make sense of the scheduler’s output	1. Who is doing what? 2. Are agents accomplishing their navigation tasks? 3. Are agents maximizing their science tasks?

4.3 Design Requirements

Phase 1 identified that worldview debugging requires considerable cognitive effort, supporting our decision to investigate how a visual analytics tool can support these sets of tasks. In phase 2, over months of iterative co-design with P0, we elaborated on the specific challenges researchers face when debugging MRS systems (summarized in Table 3). We use these questions to devise a set of requirements on how a visual analytics system can provide the needed support:

- R1. Display worldview synchronization state:** Determining if the worldviews are in sync is the first step to an in-depth analysis. The system’s visual encoding should indicate whether agents are in sync and highlight those who are not.
- R2. Support system performance assessment:** System performance is based on the number of accomplished tasks. With n agents, the number of tasks and task transferring can be high. To track who is doing what, the system should (i) differentiate the different tasks agents can perform and (ii) offer the flexibility to switch analytical perspectives (i.e., AC, TC).
- R3. Show the differences and similarities of the worldviews:** When worldview desynchronization occurs (R1), researchers will need to detect who is out of sync in order to plan for corrective action. Detecting out of sync agents requires understanding the similarities and differences of all the worldviews.
- R4. Help conduct a root cause analysis of the desynchronized worldviews:** After determining which agents are out of sync (R3), operators will need to conduct a root cause analysis to reason on why this desynchronization occurred and plan for corrective action. The system should provide the ability for users to collate and fuse pieces of information in order to properly diagnose the desynchronization condition.

5 APPLICATION

To support analyzing and debugging MRS, we design a system composed of two components: Main View (Fig. 2) and Differential Worldview Comparison (DWC) (Fig. 4). The two components are laid side-by-side, and users can first use the Main View to assess whether agents’ worldviews are synchronized. If they conclude worldviews are synchronized, they can proceed to assess the system performance with the various views displayed in the Main View. Otherwise, users can use DWC to identify worldview differences. After identifying and selecting misbehaving agents in DWC, users can engage with the Main View to conduct a root cause analysis.

5.1 Main View

The Main View (Fig. 2) consists of a summary overview, scatterplot, graph, task abstraction, and a timeline.

Summary Overview. The summary overview (Fig. 2a) provides a quick assessment of agents’ performance of science tasks and worldview synchronization (R1, R2). As navigation tasks are required, the operator can infer how well the MOSAIC scheduler is performing by assessing the number of optional science tasks that agents can accomplish. Thus, for each task in the chain of science tasks, we display how many agents have executed the respective task as a fraction of all eligible agents. In Fig. 2a, 2 out of 4 agents have accomplished the first two science tasks. Of the two, only one has accomplished the third science task (either on board or by delegating

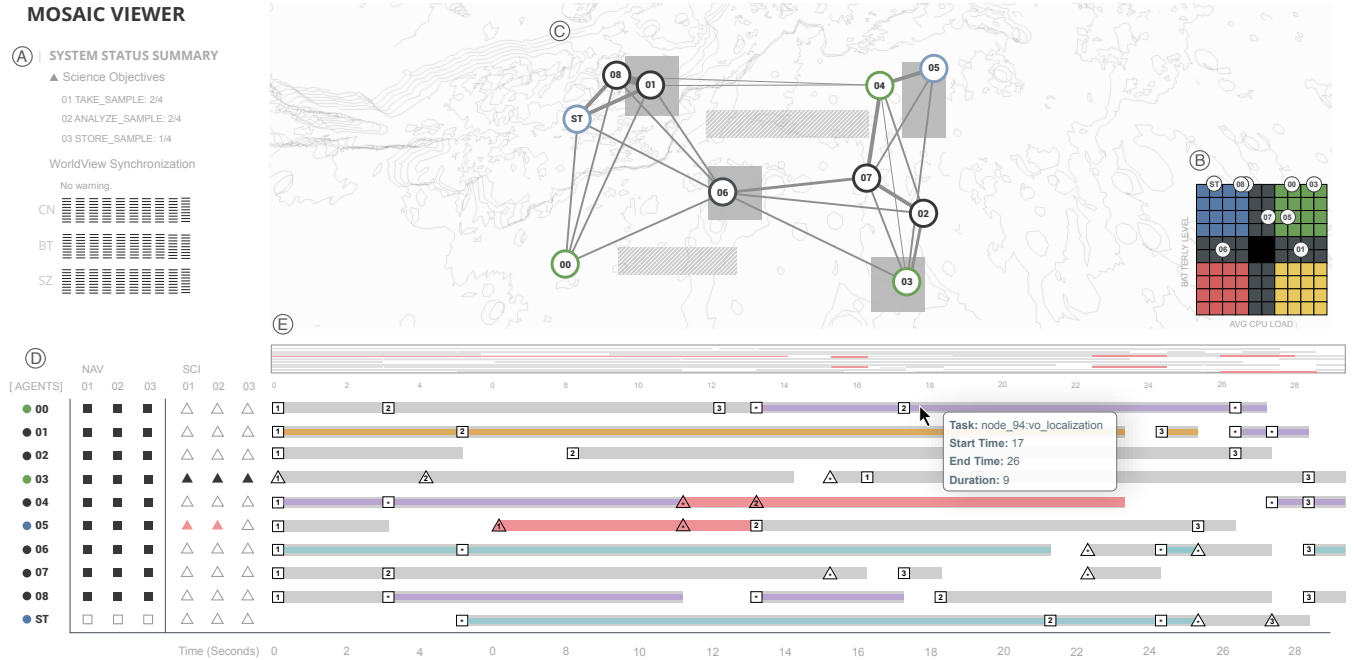


Figure 2: The operator is using the Main View to evaluate the system's performance. (a) Summary overview shows state of the science objective and the worldview synchronization (b) Scatterplot abstracts the “behavior” of agents with the x- and y-axis encoding average CPU load and battery level, respectively. (c) Graph depicts the agents' location and communication links (d) Task Abstraction provides a task-centric perspective of each agent's task. Squares represent navigation tasks. Triangles represent science tasks (e) Timeline shows agents' activities. Agent 5's science chain of task is highlighted in pink. (Note: other agents' timelines are highlighted in different colors for illustrative purposes for the case study in Sect. 6.3)

it). Below, *Worldview Synchronization* provides an overview of the agents' worldview synchronization. To help users determine worldview synchronization, if the system detects a desynchronization, it outputs a warning (R1). Users can then turn to DWC for further analysis. To show *what* is desynchronized, the Worldview Synchronization displays DWC's visual representation for three worldview attributes ('CN': Communication Network, 'BT': Battery Level, and 'SZ': Science Zone). We discuss this further in Sect. 5.2.

Scatterplot. The scatterplot (Fig. 2b) is broken into four quadrants and the x- and y-axis shows average CPU load and battery level, respectively. The quadrants are colored with four different categorical colors to help operators abstract agents' behavior at-a-glance with respect to the current time step (R2). CPU load, in particular, shows how busy a given agent is. This information in addition to battery level provides two indications: (1) it identifies over-subscribed agents and “bottlenecks”; (2) it identifies agents that are doing the majority of the work for the overall system. For example, agents in the blue quadrant are considered “lazy”. With a high battery level and low average CPU load, this suggests that these agents could be given more tasks if the communication topology allowed for it. In contrast, agents in the yellow quadrant have a low battery level and high average CPU load—indicating they are “over-worked”. The middle portion of the scatterplot is colored grey, with the center colored black, in order to emphasize extreme behavior.

Graph. In the graph (Fig. 2c), each agent is represented as a circle with their agent number in the center. The base station is denoted as ‘ST’. The colored ring correlates to which quadrant an agent is positioned in the scatterplot. The edges between agents represent its communication links, and the edge weight encodes the available communication bandwidth. The graph has two regions to represent the agent's physical environment. Solid dark grey filled-in regions are the ‘science zones’, while the diagonal hashed regions represent ‘communication cut-off zones’. Communication links that cross a communication cut-off zone are severed, simulating the effect of line-of-sight of obstructions (e.g., hills). The graph provides basic interactions to zoom-in, pan, and see details on-demand (i.e., tooltip)

when hovering over the nodes or edges. Users can also click on an agent and see its immediate edges in pink, while the opacity of the other edges in the network is lowered.

Task Abstraction and Timeline. In the Task Abstraction (Fig. 2d), each row represents an agent, and a colored circle next to the agent name corresponds to the quadrant the agent is respectively placed in the scatterplot. To help distinguish tasks, there are two sets of shapes, where squares represent navigation tasks and triangles represent science tasks (R2). Each chain of tasks is composed of three steps, and for every agent, the respective shape is filled in if the task has been accomplished. This allows operators to quickly assess if agents have accomplished their navigation tasks, and how many science tasks have been performed from a TC perspective (R2). However, this task abstraction does not tell *which* agent has completed the task or *when* it has been accomplished. To that end, operators can use the timeline, which shows each individual agent's activity (Fig. 2e) as horizontal bars in seconds. The length of a bar represents the duration of a task and the positioning maps when a task has started and ended. The shape at the beginning of each task represents the task type (navigation or science), and inside each shape is either a number that represents the nth-step of its respective chain of tasks or an asterisk symbol (*) to denote task relocation. At first sight, it is not evidently clear if a task in an agent's timeline may have been relocated from another agent for assistance (e.g., Agent 0 is performing Agent 4's second navigation task at $t = 17$). To help provide a closer inspection, users can brush the range on the timeline with a mini-timeline or see more details of a task with a tooltip. The tooltip displays the name, the start and end time of the task, and who the task belongs to. More interactions are listed in Sect. 5.3.

5.2 Differential WorldView Comparison

Differential Worldview Comparison (DWC) enacts the concept of the *diff* algorithm for text comparison [35] to compare the agents' worldviews. For each line in two text files, the diff algorithm either generates nothing, where the two lines are the same, or shows a side-by-side comparison of the two lines, where they differ. Analogously,

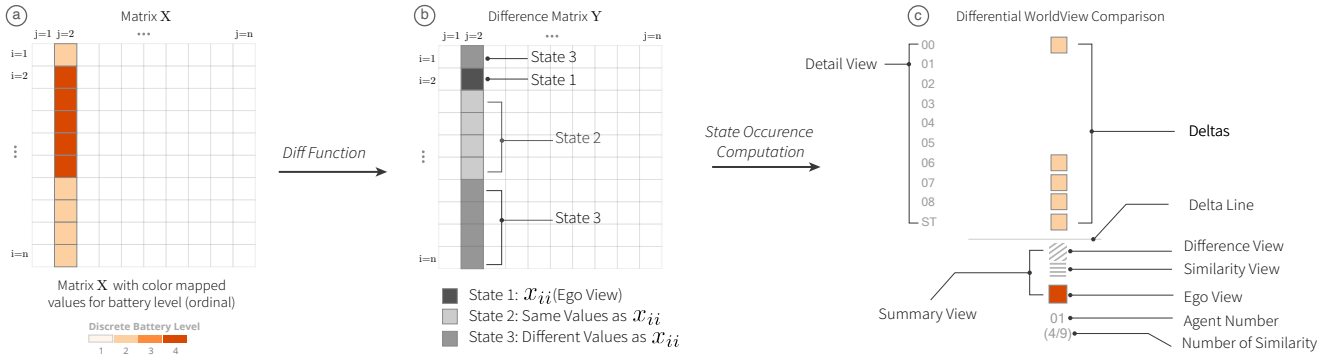


Figure 3: The process from the raw data to DWC. This figure focuses on Agent 2's battery level (Fig. 4). (a) The agents' view of a single attribute is represented as \mathbf{X} , a $n \times n$ matrix. After applying our variant of the diff function to \mathbf{X} , we obtain the Difference Matrix \mathbf{Y} (b) where each entry is categorized as either State 1, 2, or 3. We compute the number of occurrences for each state. (c) The information is transformed into DWC.

WORLDVIEW COMPARISON

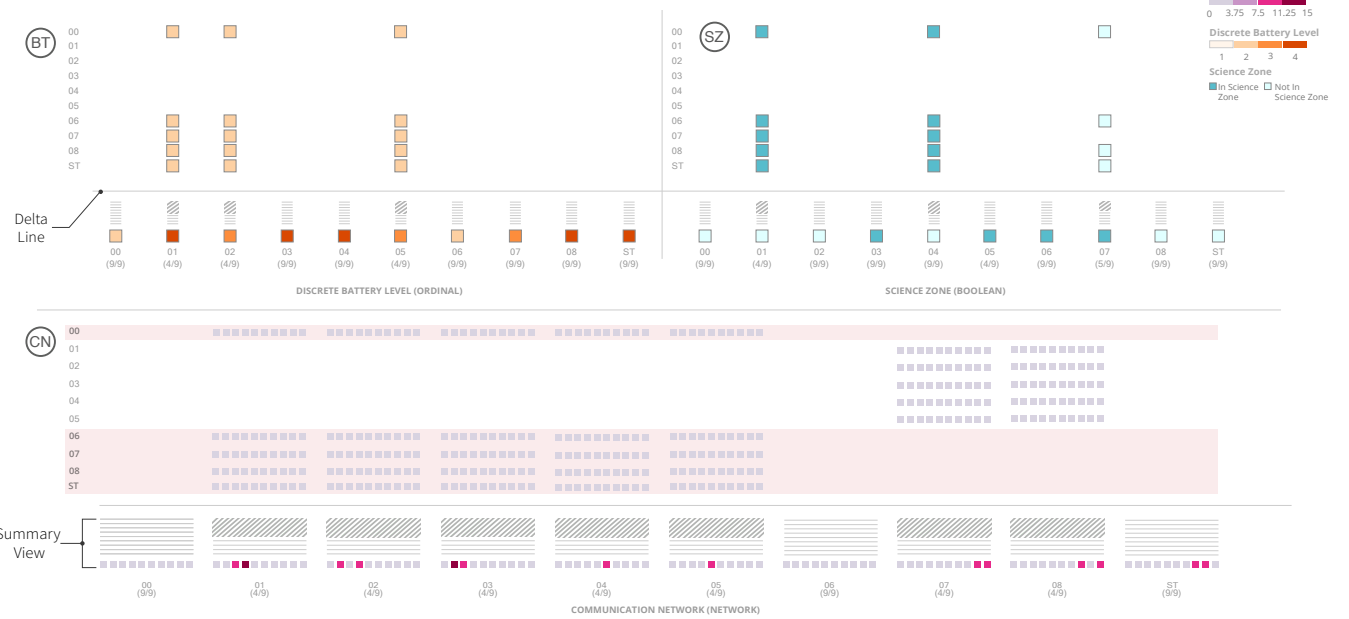


Figure 4: Differential Worldview Comparison: (BT) Battery Level Panel, (SZ) Science Zone Panel, and (CN) Communication Network Panel.

for each attribute in the agents' worldviews, we introduce a variant of the diff function that compares each agent's presumed value for an agent's attribute (e.g., agent B's presumed value for agent A's location) with the ego value of the said attribute (i.e., the value the respective agent has determined for itself. See Table 1). We compare to an agent's ego value because we follow the strong assumption that every agent knows its state the best. Hence, the agents' ego value for each attribute acts as a form of comparison. If an agent's presumed value corresponds to the respective ego value, DWC shows nothing; however, if they differ, DWC shows the presumed value. Fig. 3 provides a walk-through of the process from the raw data to DWC.

Of the six attributes listed in Table 1, we focus on the three attributes (battery level, presence in a science zone, and communication network bandwidths) that have a *direct* impact on the agents' scheduling decisions and are able to explain the majority of the desynchronization scenarios. For each agent i , each individual world-view attribute (e.g., battery level) can be represented as a 1D-array x_i with n entries, where n represents the total number of agents (Fig. 3a). The j -th entry in the array, x_{ij} represents agent i 's belief about the state of agent j 's attribute. For battery level, the entry is an integer; for science zone, it is a boolean; and for communications, the entry is a list of bandwidths from agent j to all other agents.

Once all n arrays are concatenated, this becomes an $n \times n$ matrix \mathbf{X} . In \mathbf{X} , row i represents agent i 's beliefs about the attribute, and

the entry x_{ij} represents agent i 's belief about the state of agent j 's attribute value. The value x_{ii} is denoted as the *ego value* for the attribute, as this entry represents what agent i thinks about itself. For example, the highlighted column in Fig. 3a represents every agent's belief about the state of Agent 2's battery level (encoded by color), and x_{22} represents Agent 2's ego value for its battery level.

After compiling \mathbf{X} for each attribute, we apply our variant of the *diff* function to every column j in \mathbf{X} , comparing every entry (x_{ij}) to the ego value:

$$\text{diff}(x_{ii}, x_{ij}) = \begin{cases} \text{None,} & \text{if } x_{ii} = x_{ij} \\ x_{ij}, & \text{otherwise} \end{cases} \quad (1)$$

If $x_{ii} = x_{ij}$, the function does not return anything; otherwise the value is x_{ij} (i.e., agent i 's presumed value for agent j). Once we compute the diff function for every column, the output values are used to create the Difference Matrix \mathbf{Y} (Fig. 3b). In \mathbf{Y} , each entry is categorized as one of three states: State 1, State 2, or State 3. State 1 is the ego view (y_{ii}), (ii) State 2 are entries ($y_{ij} = \text{None}$) that agree with the ego view, and (iii) State 3 are those ($y_{ij} \neq \text{None}$) that disagree with the ego view. Then, for every column j in \mathbf{Y} , we compute the Similarity and Difference Sum to count the number of entries labeled as State 2 and State 3, respectively:

$$\text{Similarity Sum} = \sum_{i=1}^n [y_{ij} = \text{State 2}], \quad (2)$$

$$\text{Difference Sum} = \sum_{i=1}^n [y_{ij} = \text{State } 3] . \quad (3)$$

We now have all the information we need to create DWC.

DWC displays each attribute as a grid-like panel (Fig. 4). Every panel has n adjacent columns, each representing beliefs about an agent. Each column j is composed of the same four components that make up DWC: Ego View, Similarity View, Difference View, and Deltas. Every panel also contains a ‘Delta Line’. Below the Delta Line, the bottom portion of every panel is the ‘Summary View’ that summarizes the system’s synchronization state, and the bottom-most component is the ‘Ego View’, which visualizes an agent’s ego value.

The visual representation of the ego value is specific to each attribute. For Communication Network, the ego view of each agent’s communication network (i.e., the bandwidths from the agent to all other agents) is represented as a 1D-array of length n . The k -th entry in the array represents the communication bandwidth value from agent j to agent k ; we encode the bandwidth value with a purple-to-red sequential colormap [31]. In contrast, the data types used in the Battery Level Panel and Science Zone Panel are ordinal and boolean, respectively; for these, we use a square mark to represent the Ego View. Ordinal values are encoded with an orange sequential colormap, and boolean values are encoded with two shades of teal.

Next, above the Ego View, the ‘Similarity View’ displays the similarities of each agent’s worldview through a piling metaphor [5] where the number of horizontal lines represents the Similarity Sum (Equation 2). This information is also captured as a fraction below the Ego View. The piling metaphor fits the need of the visualization as it visually aggregates information—reducing visual noises and allowing more emphasis on the differences. This visual design decision is also based on what we have learned from the formative study, as researchers are more interested in identifying the differences of agents’ worldviews rather than the similarities.

The next component of DWC is the ‘Difference View’, which is adjacent to the Similarity View. The Difference View is represented by diagonal hashed lines, and it complements the Similarity View. From the Difference Sum (Equation 3), the height of the Difference View indicates how many agents disagree with an agent’s ego value (R1). To see who the contrarians are and their beliefs, we look at the top portion of DWC: the Detail View (R1, R3). Above the Delta Line, the Detail View has n rows representing the n agents. For each column j , rows corresponding to an agent that disagrees with the column’s ego view (i.e., rows corresponding to an entry in State 3 in **Y**) report the contrarian agent’s belief about agent j ’s worldview. For example, in Fig. 4 (BT), Agent 7 disagrees with Agent 1 and 4’s ego view of their location. The Detail View displays a different color compared to the respective Ego View’s at the bottom. By default, rows corresponding to agents that agree with the Ego View are blank, in line with the diff metaphor. However, users can also show values that correspond to the Ego View by toggling the Similarity View.

Our decision to focus on representing the agents’ view of each attribute as a single matrix is based on the lessons from past designs from the co-design sessions. The past designs layered various visual encoding for each attribute, and users found it difficult to make sense of the layered result and to compare the differences between the agent’s worldviews. In addition, we found that operators would focus on a particular attribute depending on the context of the problem.

5.3 User Interactions

The prototype features a rich set of user interactions:

Highlighting. To support switching analytical perspectives (TC, AC) when making sense of an agent’s schedule, users can highlight a single task or highlight an agent’s chain of tasks (R2). In Fig. 2, the user highlighted Agent 5’s science chain of tasks. Users can also highlight rows in the Detail View in the DWC as shown in Fig. 4.

Interlink Views. Visualizations for MRS can be broken down into two aspects: (i) the behavior of a single agent or (ii) the overall

behavior of the system [61]. No single tool is capable of providing a complete picture of the system [49]. Hence, we focus on providing the operator with the ability to collate and fuse pieces of information in order to properly diagnose the system by interlinking the views (R4). For instance, when the rows in the Communication Network panel are highlighted, the immediate edges of the highlighted agents are pink in the graph, while the opacity of the other edges are lowered. Fig. 8 showcase this interaction. As another example, the graph, scatterplot, and task abstraction are interlinked. When an agent is selected in any of these views, it is simultaneously highlighted in the other views.

5.4 System Architecture & Implementation

We use a MongoDB database to store each agent’s reported worldview. From the database, we perform three computational tasks before visualizing the system: (i) compute summary statistics; (ii) chain tasks according to their precedence constraints and inter-task dependencies; (iii) compare agents’ worldview through the diff function (Sect. 5.2). Afterward, we visualize the results on a web application. The front-end is implemented with a combination of HTML5, CSS, JavaScript, and the JavaScript Data-Driven Documents (D3) library [9]. The back-end runs on a Node.js web server.

6 SYSTEM EVALUATION

We are interested in evaluating if MOSAIC Viewer can help operators successfully answer the critical questions listed in Table 3. We perform two assessments. First, we conduct a qualitative study that consists of a training task scenario and three task scenarios in which data were collected. For each scenario, participants were asked to explore the state of a multi-robot system. Then, based on the user study, we present two case studies that demonstrate our system’s efficacy and the workflow of how users interacted with the system.

6.1 User Study

Participants. We recruited 12 participants (3 female, 9 male), aged reported in bins 18 - 44 years, from the population of multi-robot systems researchers and operators at NASA JPL. Of the 12 participants, two were part of the formative evaluation. We elected to include these participants in the study as the system’s interactions and visual representations changed after formative evaluation, and the study was designed to assess their ability to debug worldviews based on scenarios they did not encounter in the past. A more comprehensive overview of the participants can be found in Table 2.

Conditions and Tasks Design. To study the usability and outcomes of domain experts using MOSAIC Viewer, we devised three scenarios, each touching upon a common situation operators face. One scenario included all agents in sync. Two scenarios included an “out of sync” condition that emerges from (i) one agent isolated from others or (ii) a bipartition in the agents’ communication network, respectively. For each scenario, we asked operators to assess the system. If they determined worldviews are out of sync, we asked operators to perform a root cause analysis. This involved determining which robots were out of sync and analyzing why they were out of sync (Q1). Participants were also asked to assess whether agents accomplished their science tasks (Q2), navigation tasks (Q3) as well as whether they understood how tasks were scheduled (Q4). Each participant was required to complete the three scenarios, and the study was counterbalanced to mitigate learning effects. We chose not to have a baseline interface to compare with MOSAIC Viewer instead of using existing debugging tools for two reasons: (1) reduce confounding effects that may emerge from other aspects of the interfaces and (2) focus the investigation on the qualitative, behavioral aspects participants gained from MOSAIC Viewer.

Experimental Setup. MOSAIC Viewer ran on a mid-2017 MacBook Pro (16 GB, 2.5 GHz process). The interface was displayed on a 34-inch display (3440 × 1440 pixels), using the Google Chrome

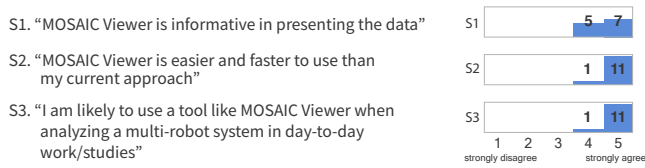


Figure 5: Participant's feedback about MOSAIC Viewer on a 5-point Likert Scale. Median ratings are indicated in gray.

browser. Participant input was captured through an external keyboard and mouse. For the sake of uniformity, pen and paper were provided to participants regardless of the task.

Procedure. Each participant first filled in their background information in a survey form. They were then trained on the interface until they were comfortable using it. Once ready, we provided the three scenarios one by one. For each scenario, the participants were asked to answer the four aforementioned questions and write down their answers on the provided sheet of paper. At the end of each task, participants responded to a questionnaire about the confidence of their answers. After finishing all three tasks, participants answered additional survey questions that prompted their overall thoughts about the system and engaged in a semi-structured exit interview. We used a concurrent think-aloud protocol during the study, and participants were audio- and screen-recorded for the duration of the tasks. All of the participant's answers for the tasks were saved.

6.2 Results

Our evaluation revealed the following findings:

- **Speed and interactivity streamline higher-level analyses:** (1) MOSAIC Viewer helps users locate information faster; (2) The interlinked views enable users to formulate a hypothesis about the root cause of the desynchronized worldviews.
- **Trust for summary displays grew with experience:** Users initially lacked trust in the visual representations to which they were not accustomed, but this trust grew quickly once they were able to verify their understanding.
- **The why: The how is not enough on its own:** Understanding MRS agents requires not only understanding how agents are interacting with each other and how the tasks are scheduled but also why. One without the other is not complete.
- **Different sets of assumptions affect data interpretation:** Users bring in a different set of assumptions that mismatches from the system's architecture. These mismatches led to incorrect data interpretation.

P6 explains that MOSAIC Viewer meets an unmet need with existing tools, noting "Every time I build a new capability...There are no tools out of the box that just does it for you. You have to go spend time and build it, so you can properly visualize what our algorithms do."

6.2.1 Speed and interactivity streamline analyses

Navigating information. All 12 users reported that understanding the Main View's visual encoding required very low to low-level of effort. P7 attributes the high learnability due to the fact "someone actually thought about how to represent these data rather than [users] just plotting the data in a given software". In particular, participants reported the graph and shapes to be intuitive and useful.

8 out of 12 participants helped explain this finding, sharing that their current workflows required them to open multiple terminal windows per agent, often spreading across multiple monitors to unpack what even a single agent is doing. This matches the finding from the formative study. Even when working alone, participants describe terminal logs as time-consuming and inefficient, especially for comparisons. P1 explains this workflow requires visually scanning and remembering the contents—making it mentally taxing. P2 contrasts this multi-screen collaboration experience with MOSAIC Viewer's



Figure 6: Toggle interaction used in the DWC. (a) show the default setting, only showing the deltas in the Detail View. (b) shows when P4 toggles the Similarity View and the agents that agree with Agent 1's ego view of its battery level appears in the Detail View. (c) shows when P4 toggles the Difference View and makes the deltas disappear.

compact form, observing that "[in MOSAIC Viewer] whatever I want to see, the information is there."

Users contrasted MOSAIC Viewer with terminal logs, which require considerable work from users to find individual pieces of information, let alone combine them into a higher-level analysis. P1 comments, "that isn't possible with the way I [currently] do it". Based on the feedback users provided on a 5-point Likert scale, participants describe MOSAIC Viewer as both easier- and faster-to-use than their current approaches ($Md = 5$, $IQR = 0$, $min = 1$) (Fig. 5). Others found the flexibility helped drive ease-of-use, with P2 observing "you have more than one way to see [the same] information".

Formulating. Previous research explains that debugging a distributed multi-robot system requires users to combine macro- (i.e., societal-level) and micro-level (i.e., agent-level) system state information to form a coherent, unified picture [49]. In our study, users explain that the speed of specific data access enables them to more quickly achieve these higher levels of understanding. To complete each task, users would use the DWC to identify if an agent is out of sync with the other agents. To explain a de-synchronization, users would have to combine their domain knowledge and the information provided from each of the various views to conduct a root cause analysis (R4). Users would look at each view to determine if the agent's location, distance, communication bandwidth, or any other variables could explain the desynchronization they determined in DWC. In Sect. 6.3, we provide a walk-through of how users utilized our system with two case studies.

6.2.2 Trust for summary displays grew with experience

Independent of which task was presented first, users first interacted with the system in a way that indicated they were verifying the low-level details in which the summary displays were based upon.

The DWC Summary View provides a good example of this observed user behavior. The default setting of DWC shows only the differences in the Detail View, but users can toggle different parts of the Summary View to see different low-level details (Fig. 6). Another example is the Task Abstraction view. The Task Abstraction view utilizes filled-in shapes to represent the three step process of the science and navigation chain of tasks; however the shapes do not indicate when or who has executed the respective task. To verify the abstraction, users can individually click on a shape and see the particular task being highlighted in the timeline visualization or automatically highlight the entire chain of tasks as shown in Fig. 4.

Fig. 7 shows the average number of clicks users used to either highlight and toggle throughout the user study's three tasks. In the first task, 8 of 12 users used the toggle interaction to validate their understanding of the DWC's visual encoding (Fig. 6). As P4 explains: "The reason why I've been [toggling] is because I'm not sure right now whether I'm seeing the similarity or differences (in the Detail View). The way I can check that is by looking at the

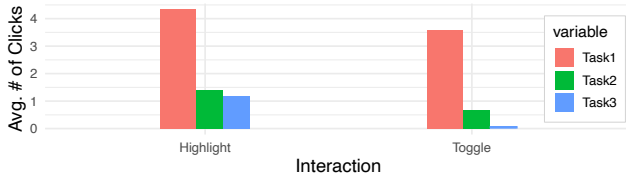


Figure 7: Summary of participants' average number of clicks for two interactions. Participants' average number of clicks decreases over the study. The average for the toggle interaction in the third task is 0.

color of [Agent 1's ego view of battery level], and toggling the [diagonal lines] to confirm what I'm seeing." By the second task, only 3 users continued to use the toggle interaction to validate their understanding, and 0 users used the toggle interaction by the third task. As P6 explains, "At first, I [toggled] because I wanted to see the whole picture. But once I got used to the system, I don't need to verify. I know from here it's going to show the same thing so once you get used to it, it's not necessary."

A similar pattern of use was observed when users were asked whether agents have completed their science objectives. During the first task, 7 out of 12 users used the chain of task highlighting interaction to confirm whether an agent's science task abstraction was accurately reflected in the timeline visualization. By the third task, only 3 used the interaction, while the rest used the Task Abstraction exclusively. P8 commented: "It took me a while to really (pause) trust that the little [shapes] were fully representative (of the timeline visualization). I just wanted to double-check that's the right answer."

6.2.3 What and how are not enough: expert users need to know why

While participants shared how MOSAIC Viewer successfully explains what the agents are doing and provide insight to some anomalous behaviors, expert operators wanted even more detail than the system provided. P6 provides a representative quote:

Usually, people using [the system] not only want to see how it's scheduled. I think [answering] how [the tasks] are scheduled, the visualizations are doing that beautifully. But there's some underlying optimization going on where the objective function is trying to maximize something...As a human operating a system, one of the things we always want to do is verification. So my computer is telling me this [solution] is the best. Is it really the best? Can I check the **sanity** of the solution?

Similarly, 10 participants indicated they would like to know more about the reasoning behind the scheduling optimization with comments such as, "I feel like 7 and [base station] should be doing better than that and I wonder why they're not", "I want to know why [Agent 5's science task] was transferred to 4. Why not to 0 or 2?", "Why are there blank spaces in the timeline?". During the in-depth interviews, participants commented that they would like to see explanations behind the scheduler's decisions. For example, when an agent's task is delegated to another agent, P6 and P10 explain they would want to see the low-level details, such as real-time CPU load, of the two agents in order to understand the scheduling optimization.

6.2.4 Different sets of assumptions affect sensemaking

Despite an unbounded training period, participants who are not familiar with MOSAIC's specific logistics displayed behavior where they interpreted data differently. We note two specific cases.

P7 and P9, two non-core MOSAIC participants, incorrectly answered "no" for all three tasks when asked if the agents accomplished their navigation task (Q3). In their think-aloud process, both mentioned the squares that represent the base station's navigation chain of task in the Task Abstraction are not filled in, indicating that the base station failed to accomplish its navigation tasks. However,

as explained in Sect. 2.2, the base station is a special agent that does not need to accomplish the mandatory navigation task as its purpose is to help other agents with its fast computation power.

A similar pattern of mental model mismatch occurred with P4. During the first task, P4 provided an incorrect answer to the question related to the agents' science tasks (Q2). In their think-aloud process, P4 observed how Agent 3 accomplished its science chain of task with the chain of task highlighting and stated the answer to Q2 is "yes". However, according to the Task Abstraction in Fig. 2, Agent 5 did not accomplish its science task. From the exit interview, P4 elaborated the reasoning behind their answer is based on the idea of how "the agents are working together in tandem" where if one agent accomplished its science chain of tasks, other agents did as well.

6.3 Case Studies

Evaluate System Performance. Upon launching the system, the operator sees the Main View and DWC side-by-side. The Summary View (Fig. 2a) does not display a synchronization warning and DWC also does not show any deltas. The operator concludes the agents' worldviews are in sync and moves on to assess the system's performance (i.e., how many tasks are being performed, and by whom) as part of their next goal (Table 3).

The operator looks at the Task Abstraction (Fig. 2d) to obtain a high-level overview of the system's performance. The filled-in squares in the "NAV" column indicate agents have accomplished their mandatory navigation task. The "SCI" column shows Agent 3, with its three filled-in triangles, is the only agent that fully accomplished its science chain of tasks. However, the Science Objectives (Fig. 2a) and the graph (Fig. 2c) show there are a total of four eligible agents (Agents 1, 3, 5, and 6) that can accomplish science tasks.

The operator engages with the timeline (Fig. 2e) and scatterplot (Fig. 2b) to seek for the low-level details that explain why the other three agents failed to accomplish their science tasks. Highlighting Agent 5's partial science chain of tasks (Fig. 2e colored in pink), the operator sees Agent 5 relocated the second science task to Agent 4. From the graph, the shortest path to send the scientific data from Agent 4 to the base station is through Agent 8 or Agent 1. However, Agent 4 has a weak communication link to both agents and is unable to relay the data within the time gap from $t = 23$ to $t = 27$. Next, the scatterplot shows that Agent 1's and Agent 6's battery level is 50%, indicating the computation time needed for their tasks will take longer compared to other agents (e.g., Agent 1 takes twice as long as Agent 0 to accomplish its second navigation task). Evidently, Agent 1 and Agent 6 (colored in orange and turquoise, respectively) did not accomplish any science tasks as both needed more time to accomplish their mandatory navigation tasks.

Plan for Corrective Action. With the same "out of sync" scenario from the user study, this case study reflects how operators can plan for corrective action (i.e., determine who is out of sync and reason why the desynchronization occurred) with MOSAIC Viewer.

Launching the system, the operator sees the Main View and DWC side-by-side. The Main View signals a desynchronization warning in the Summary Overview, and DWC (Fig. 4) displays deltas across all three panels. The deltas specifically in the Communication Network Panel (Fig. 8b) show two distinct sets. Highlighting rows 0, 6, 7, 8, ST in the Communication Network panel emphasizes the visualization's white space, revealing a bipartition in the communication network as Agents 0, 6, 7, 8, ST are out of sync with Agents 1-5, and conversely Agents 1-5 are out of sync with Agent 7 and 8. The deltas in the other two panels have the same visual pattern and the same set of agents (Agents 0, 6, 7, 8, and ST) are out of sync.

The pattern revealed by the DWC is also displayed in the graph in the Main View (Fig. 8a): the same grouping of agents in DWC is reflected within the two distinct clusters of the graph. The weak communication bandwidth between the two clusters (indicated by black arrows) further supports the hypothesis of a bipartite graph.

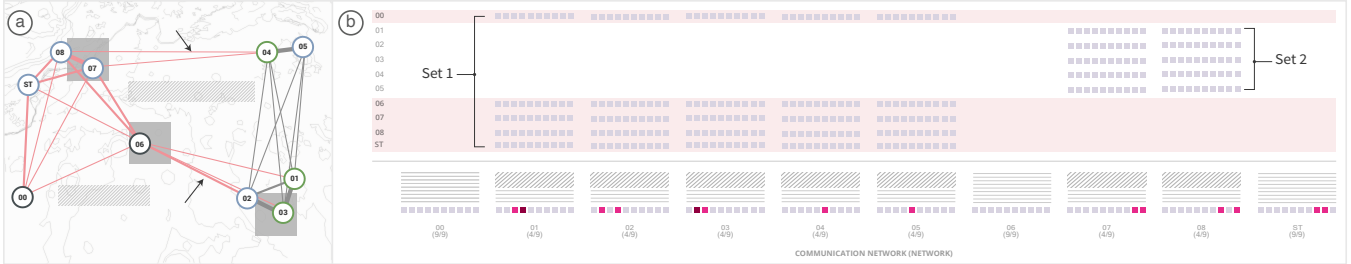


Figure 8: (a) Graph and (b) Communication Network Panel. In (b), operator highlights row 0, 6, 7, 8, ST and identifies a visual pattern of how Set 1 is the complement set of Set 2. They hypothesize there’s a network bipartition. The immediate edges of these selected agents are highlighted in (a) and the same group of agents are separated. The weak communication links (indicated by black arrows) further support this hypothesis.

For instance, the bandwidth value between Agent 4 and 7 is 1, a weak value where agents would not be able to send data across the network and update each others’ worldviews, causing the bipartition.

7 DISCUSSION

In this section, we reflect upon the findings of this work, and consider both its extension, and its limitations.

Limitations in Visual Scalability. While the scale of ten agents in our work aligns with current and proposed space exploration concepts [8, 19, 20, 45] within the next 10 years, the current system design might struggle to achieve the same levels of legibility and low cognitive load with swarm robotics [48] mission concepts that include hundreds of agents. Edge bundling [24] and layout algorithms [41] are promising directions to improve dense graph legibility and transparency [13], and jitter [71] can improve dense scatterplot legibility. However, critically, in swarm robotics applications, agents are rarely endowed with a global worldview—rather, each agent typically relies on much simpler scheduling schemes and reacts based on its own state and immediate neighbors’. Accordingly, we expect that further research will be required to capture the qualitatively different nature of the underlying problem of debugging autonomous swarms’ behaviors.

Foreground scheduler logistics to achieve generalizability. In our findings, we observed differences between how core MOSAIC and non-core researchers interpreted the same visual marks on the display. In the example reported in Sect 6.2.4, the Task Abstraction shows unfilled squares for the base station. Some non-core team members interpreted that not all robots have completed their navigation tasks when seeing unfilled squares for the base station. However, core team members saw the same marks, and, knowing that the base station does not need to complete the navigation task, made the correct interpretation that all mandatory navigation tasks were completed. Reflecting on the systems’ visual encoding, we realize this confusion could have been avoided by adding an additional glyph that distinguishes the base station from other agents.

This split in our test population reveals an interesting artifact of our design process. Though we spent nearly a year on iteratively designing, we never observed this problem in our user testing until we included users outside of the core MOSAIC team. We reflect that while both groups contain full-time MRS robotics researchers, the core group is more deeply steeped in the architecture and algorithms that drive the MOSAIC scheduler. To design a more general MRS application that would work across any number of scheduling algorithms, this would require to dive deep into the mechanics of scheduling algorithms. A general system would need to visualize not only the various states of the worldviews, but also many of the underlying mechanisms through which the scheduling system operates. With this approach, designers could be more likely to create a system that functions independently of the background of its users.

Move Beyond the Debugging Use Case. This work focuses more explicitly on explaining *what* the autonomy has decided. Given this success of this work, participants also expressed some desire to understand, as P7 states, “different perspectives of the same

code”, that extends to understanding *why* the autonomy made the decisions it did in non-error cases. This points towards supporting the task of scheduling algorithm design, such as adding a “what-if” mode [76], to help roboticists gain insight into tasks that include how different algorithm hyperparameters affect the search of a very large space. This creates an opportunity to work in algorithm visualization [28, 64], as well as large decisions trees visualization [67, 70]

Extend to other Multi-Robot Systems. While this system was designed to support users of the MOSAIC scheduler, it also represents an interesting challenge to extend the application to other MRS systems that use different underlying algorithms and for different contexts of use. In one exit interview, for example, P5 observed that MOSAIC Viewer might also support agent navigation discrepancies in support of DARPA’s Subterranean Challenge (SubT) [15]. P5 reflected on ways in which the contexts share many commonalities that could help the systems achieve more generalizability, noting “What you’re visualizing is at the core of these multi-robot problems. They are connected by some network infrastructure, they’re sharing information in order to come to a decision about what to do. So we want to visualize what were the states at a specific time, what were they doing, what were they planning on doing, and how they were connected with each other.” But they also noted other system differences that might not make the contexts topologically equivalent, noting “[SubT’s] task network is a bit larger [than MOSAIC’s] and not as straightforward as 1, 2, 3. You roll back and there’s loops. So it’s harder to just lay it out sequentially like [MOSAIC’s]”.

8 CONCLUSION

We present MOSAIC Viewer, a visual analytics system that helps users make sense of robots’ autonomous scheduling decisions and pinpoint the cause of desynchronized worldviews in a multi-robot system. To compare worldviews, we draw inspiration from the *diff* algorithm to visually emphasize the differences, while aggregating the similarities. This approach allows users to quickly detect the differences and similarities of all the robots’ worldviews. The interlinked views help users not only collate and fuse pieces of information from each view in order to conduct a root cause analysis of the desynchronized worldviews, but also understand the behavior of the system at a societal and individual level. Our qualitative user study with domain experts at the NASA JPL characterizes and elaborates the usefulness and effectiveness of MOSAIC Viewer.

9 ACKNOWLEDGEMENT

The development of MOSAIC Viewer was enabled by the JPL/Caltech/ArtCenter data visualization program. We would like to thank Santiago Lombeyda, Hillary Mushkin, Maggie Hendrie, Alessandra Fleck, and Sarah Strickler for their feedback and contribution on the earlier prototypes of MOSAIC Viewer. Also, special thanks to the MOSAIC team! This research is sponsored in part by the U.S. National Science Foundation through grant IIS-1741536. The research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration (80NM0018D0004).

REFERENCES

- [1] D. Altheide, M. Coyle, K. DeVriese, and C. Schneider. Emergent qualitative document analysis. *Handbook of Emergent Methods*, pp. 127–151, 2008.
- [2] B. Annable, D. Budden, and A. Mendes. Nubugger: A visual real-time robot debugging system. In *Proc. RoboCup*, pp. 544–551. Springer, 2013.
- [3] T. Arai, E. Pagello, L. E. Parker, et al. Advances in multi-robot systems. *IEEE Transactions on Robotics and Automation*, 18(5):655–661, 2002.
- [4] B. Argrow, D. Lawrence, and E. Rasmussen. UAV systems for sensor dispersal, telemetry, and visualization in hazardous environments. In *Proc. AIAA Aerospace Sciences Meeting and Exhibit*, p. 1237, 2005.
- [5] B. Bach, N. Henry-Riche, T. Dwyer, T. Madhyastha, J.-D. Fekete, and T. Grabowski. Small MultiPiles: Piling time to explore temporal patterns in dynamic networks. *Computer Graphics Forum*, 34(3):31–40, 2015.
- [6] D. J. Barnes, M. T. Russell, and M. C. Wheadon. Developing and adapting UNIX tools for workstations. In *EUUG Conference Proceedings*, pp. 321–333, 1988.
- [7] J. L. Baxter, E. Burke, J. M. Garibaldi, and M. Norman. Multi-robot search and rescue: A potential field based approach. In *Autonomous Robots and Agents*, pp. 9–16. Springer, 2007.
- [8] S. S. Board, N. R. Council, et al. *Vision and voyages for planetary science in the decade 2013-2022*. National Academies Press, 2012.
- [9] M. Bostock, V. Ogievetsky, and J. Heer. D³ data-driven documents. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2301–2309, 2011.
- [10] B. Burmeister, A. Haddadi, and G. Matylis. Application of multi-agent systems in traffic and transportation. *IEE Proceedings - Software Engineering*, 144(1):51–60, 1997.
- [11] K. A. Cook and J. J. Thomas. Illuminating the path: The research and development agenda for visual analytics. Technical report, Pacific Northwest National Lab., Richland, WA (United States), 2005.
- [12] M. Cummings and P. Mitchell. Managing multiple UAVs through a timeline display. In *Proc. AIAA Info Tech*, p. 7060, 2005.
- [13] T. N. Dang, L. Wilkinson, and A. Anand. Stacking graphic elements to avoid over-plotting. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1044–1052, 2010.
- [14] M. De Rosa, J. Campbell, P. Pillai, S. Goldstein, P. Lee, and T. Mowry. Distributed watchpoints: Debugging large multi-robot systems. In *Proc. ICRA*, pp. 3723–3729. IEEE, 2007.
- [15] Defense Advanced Research Projects Agency. DARPA Subterranean (SubT) Challenge. URL: <https://www.darpa.mil/program/darpa-subterranean-challenge>.
- [16] N. Elmqvist and J. S. Yi. Patterns for visualization evaluation. *Information Visualization*, 14(3):250–269, 2015.
- [17] A. Emslie, R. Lagace, and P. Strong. Theory of the propagation of uhf radio waves in coal mine tunnels. *IEEE Transactions on Antennas and Propagation*, 23(2):192–205, 1975.
- [18] E. E. Entin and D. Serfaty. Adaptive team coordination. *Human factors*, 41(2):312–325, 1999.
- [19] eoPortal. CYGNSS (Cyclone Global Navigation Satellite System). URL: <https://directory.eoportal.org/web/eoportal/satellite-missions/c-missions/cygnss>.
- [20] eoPortal. OPAL (Orbiting Picosatellite Automatic Launcher). URL: <https://directory.eoportal.org/web/eoportal/satellite-missions/o/opal>.
- [21] J. A. Fails, A. Karlson, L. Shahamat, and B. Shneiderman. A visual interface for multivariate temporal data: Finding patterns of events across multiple histories. In *Proc. VAST*, pp. 167–174. IEEE, 2006.
- [22] J. Figueiredo, N. Lau, and A. Pereira. Multi-agent debugging and monitoring framework. *IFAC Proceedings Volumes*, 39(20):114–120, 2006.
- [23] J. Gancet, E. Motard, A. Naghsh, C. Roast, M. M. Arancon, and L. Marques. User interfaces for human robot interactions with a swarm of robots in support to firefighters. In *Proc. ICRA*, pp. 2846–2851. IEEE, 2010.
- [24] E. R. Gansner, Y. Hu, S. North, and C. Scheidegger. Multilevel agglomerative edge bundling for visualizing large graphs. In *Proc. PacificVis*, pp. 187–194. IEEE, 2011.
- [25] H. Garcia-Molina, F. Germano, and W. H. Kohler. Debugging a distributed computing system. *IEEE Transactions on Software Engineering*, (2):210–219, 1984.
- [26] M. Gleicher, D. Albers, R. Walker, I. Jusufi, C. D. Hansen, and J. C. Roberts. Visual comparison for information visualization. *Information Visualization*, 10(4):289–309, 2011.
- [27] R. Gove, J. Saxe, S. Gold, A. Long, and G. Bergamo. SEEM: a scalable visualization for comparing multiple large sets of attributes for malware analysis. In *Proc. VizSec*, pp. 72–79. IEEE, 2014.
- [28] S. Grissom, M. F. McNally, and T. Naps. Algorithm visualization in CS education: Comparing levels of student engagement. In *Proc. SoftVis*, pp. 87–94. ACM, 2003.
- [29] J. Halloran, E. Hornecker, G. Fitzpatrick, M. Weal, D. Millard, D. Michaelides, D. Cruickshank, and D. De Roure. Unfolding understandings: Co-designing ubicomp in situ, over time. In *Proc. DIS*, pp. 109–118. ACM, 2006.
- [30] J. Y. Halpern and Y. Moses. Knowledge and common knowledge in a distributed environment. *Journal of the ACM*, 37(3):549–587, 1990.
- [31] M. Harrower and C. A. Brewer. ColorBrewer.org: An online tool for selecting colour schemes for maps. *Cartographic Journal*, 40(1):27–37, 2003.
- [32] J. H. Holland. *Emergence: From chaos to order*. OUP Oxford, 2000.
- [33] K. Holtzblatt and H. Beyer. *Contextual design: Defining customer-centered systems*. Elsevier, 1997.
- [34] C. M. Humphrey, S. M. Gordon, and J. A. Adams. Visualization of multiple robots during team activities. In *Proc. HFES*, vol. 50, pp. 651–655. SAGE Publications, 2006.
- [35] J. W. Hunt and M. D. MacIlroy. *An algorithm for differential file comparison*. Bell Laboratories Murray Hill, 1976.
- [36] J. Jin, R. Sanchez, R. T. Maheswaran, and P. Szekely. Vizscript: On the creation of efficient visualizations for understanding complex multi-agent systems. In *Proc. IUI*, pp. 40–49. ACM, 2008.
- [37] J. Jo, J. Huh, J. Park, B. Kim, and J. Seo. LiveGantt: Interactively visualizing a large manufacturing schedule. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2329–2338, 2014.
- [38] G. Kardas, M. Challenger, S. Yildirim, and A. Yamuc. Design and implementation of a multiagent stock trading system. *Software: Practice and Experience*, 42(10):1247–1273, 2012.
- [39] J. T. Karras, C. L. Fuller, K. C. Carpenter, A. Buscicchio, D. McKeeby, C. J. Norman, C. E. Parcheta, I. Davydychev, and R. S. Fearing. Pop-up mars rover with textile-enhanced rigid-flex PCB body. In *Robotics and Automation (ICRA), 2017 IEEE International Conference on*, pp. 5459–5466. IEEE, 2017.
- [40] J. Kasper, J. Lazio, A. Romero-Wolf, J. Lux, and T. Neilsen. The sun radio interferometer space experiment (sunrise) mission concept. In *2019 IEEE Aerospace Conference*, pp. 1–11. IEEE, 2019.
- [41] S. Kieffer, T. Dwyer, K. Marriott, and M. Wybrow. Hola: Human-like orthogonal network layout. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):349–358, 2015.
- [42] H. Kunsei, K. S. Bialkowski, M. S. Alam, and A. M. Abbosh. Improved communications in underground mines using reconfigurable antennas. *IEEE Transactions on Antennas and Propagation*, 66(12):7505–7510, 2018.
- [43] S. Liu, D. Maljovec, B. Wang, P.-T. Bremer, and V. Pascucci. Visualizing high-dimensional data: Advances in the past decade. *IEEE Transactions on Visualization and Computer Graphics*, 23(3):1249–1268, 2016.
- [44] Y. Luo, K. Liu, and D. N. Davis. A multi-agent decision support system for stock trading. *IEEE Network*, 16(1):20–27, 2002.
- [45] A. J. Mannucci, J. Dickson, C. Duncan, and K. Hurst. GNSS geospace constellation (GGC): A CubeSat space weather mission concept. Technical report, Jet Propulsion Laboratory, California Institute of Technology, 2010.
- [46] F. Michel, J. Ferber, and A. Drogoul. Multi-agent systems and simulation: A survey from the agent community’s perspective. In *Multi-Agent Systems*, pp. 17–66. CRC Press, 2018.
- [47] K. Nagatani, Y. Okada, N. Tokunaga, S. Kiribayashi, K. Yoshida, K. Ohno, E. Takeuchi, S. Tadokoro, H. Akiyama, I. Noda, et al. Multirobot exploration for search and rescue missions: A report on map building in RoboCupRescue 2009. *Journal of Field Robotics*,

- 28(3):373–387, 2011.
- [48] I. Navarro and F. Matía. An introduction to swarm robotics. *Isrn robotics*, 2013.
 - [49] D. T. Ndumu, H. S. Nwana, L. C. Lee, and J. C. Collis. Visualising and debugging distributed multi-agent systems. In *Proc. AAMAS*, pp. 326–333, 1999.
 - [50] C. Niederer, H. Stitz, R. Hourieh, F. Grassinger, W. Aigner, and M. Streit. TACO: visualizing changes in tables over time. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):677–686, 2017.
 - [51] E. Osawa, H. Kitano, M. Asada, Y. Kuniyoshi, and I. Noda. RoboCup: the robot world cup initiative. In *Proc. ICMAS*, pp. 9–13, 1996.
 - [52] L. E. Parker. Distributed intelligence: Overview of the field and its application in multi-robot systems. In *Proc. AAAI Fall Symposium: Regarding the Intelligence in Distributed Intelligent Systems*, pp. 1–6, 2007.
 - [53] C. Pinciroli, V. Trianni, R. O’Grady, G. Pini, A. Brutschy, M. Brambilla, N. Mathews, E. Ferrante, G. Di Caro, F. Ducatelle, et al. Argos: a modular, parallel, multi-engine simulator for multi-robot systems. *Swarm intelligence*, 6(4):271–295, 2012.
 - [54] J. Preece, Y. Rogers, H. Sharp, D. Benyon, S. Holland, and T. Carey. *Human-Computer Interaction*. Addison-Wesley Longman Ltd., 1994.
 - [55] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng. ROS: An open-source robot operating system. In *Proc. ICRA Workshop on Open Source Software*, vol. 3, p. 5. Kobe, Japan, 2009.
 - [56] A. Rahmani, S. Bandyopadhyay, F. Rossi, J.-P. de la Croix, J. V. Hook, and M. T. Wolf. Space vehicle swarm exploration missions: A study of key enabling technologies and gaps. In *Proc. IAC*, 2019.
 - [57] M. Rettig. Prototyping for tiny fingers. *Communications of the ACM*, 37(4):21–27, 1994.
 - [58] F. Rossi, R. Zhang, Y. Hindy, and M. Pavone. Routing autonomous vehicles in congested transportation networks: Structural properties and coordination algorithms. *Autonomous Robots*, 42(7):1427–1442, 2018.
 - [59] H. A. Ruff and G. L. Calhoun. Human supervision of multiple autonomous vehicles. Technical report, Air Force Research Lab Wright-Patterson AFB OH Human Effectiveness Directorate, 2013.
 - [60] R. Sadana, T. Major, A. Dove, and J. Stasko. Onset: A visualization technique for large-scale binary set data. *IEEE transactions on visualization and computer graphics*, 20(12):1993–2002, 2014.
 - [61] M. Schroeder and P. Noy. Multi-agent visualisation based on multivariate data. In *Proc. AAMAS*, pp. 85–91, 2001.
 - [62] C. Seah, M. Sierhuis, and W. J. Clancey. Multi-agent modeling and simulation approach for design and analysis of mer mission operations. 2005.
 - [63] M. Sedlmair, M. Meyer, and T. Munzner. Design study methodology: Reflections from the trenches and the stacks. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2431–2440, 2012.
 - [64] C. A. Shaffer, M. L. Cooper, A. J. D. Alon, M. Akbar, M. Stewart, S. Ponce, and S. H. Edwards. Algorithm visualization: The state of the field. *ACM Transactions on Computing Education*, 10(3):1–22, 2010.
 - [65] R. Simmons, D. Apfelbaum, W. Burgard, D. Fox, M. Moors, S. Thrun, and H. Younes. Coordination for multi-robot exploration and mapping. In *Proc. AAAI/IAAI*, pp. 852–858, 2000.
 - [66] H. Song, B. Lee, B. H. Kim, and J. Seo. DiffMatrix: Matrix-based interactive visualization for comparing temporal trends. In *Proc. EuroVis (Short Papers)*, 2012.
 - [67] P. Szekely, R. Maheswaran, C. M. Rogers, and R. Sanchez. Scheduling the activities of distributed teams. 2008.
 - [68] P. Szekely, C. M. Rogers, and M. Frank. Interfaces for understanding multi-agent behavior. In *Proc. IUI*, pp. 161–166. ACM, 2001.
 - [69] A. Tanoto, J. L. Du, T. Kaulmann, and U. Witkowski. MPEG-4-based interactive visualization as an analysis tool for experiments in robotics. In *Proc. MSV*, pp. 186–192, 2006.
 - [70] G. Taylor, R. M. Jones, M. Goldstein, R. Frederiksen, and R. E. Wray. VISTA: A generic toolkit for visualizing agent behavior. In *Proc. CGF*, pp. 29–40, 2002.
 - [71] M. Trutschl, G. Grinstein, and U. Cvek. Intelligently resolving point occlusion. In *Proc. InfoVis*, pp. 131–136. IEEE, 2003.
 - [72] T. S. Tullis. *Predicting the usability of alphanumeric displays*. PhD thesis, 1984.
 - [73] M. H. Van Liedekerke and N. M. Avouris. Debugging multi-agent systems. *Information and Software Technology*, 37(2):103–112, 1995.
 - [74] J. Vander Hook, W. Seto, V. Nguyen, Z. Hasnain, L. Gallagher, T. Halpin-Chan, V. Varahamurthy, and M. Angulo. Autonomous swarms of high speed maneuvering surface vessels for the central test evaluation improvement program. In *Unmanned Systems Technology XXI*, vol. 11021, p. 110210M. International Society for Optics and Photonics, 2019.
 - [75] J. Vander Hook, T. Vaquero, F. Rossi, M. Troesch, M. S. Net, J. Schoolcraft, J.-P. de la Croix, and S. Chien. Mars on-site shared analytics information and computing. In *Proc. ICAPS*, vol. 29, pp. 707–715, 2019.
 - [76] J. Wexler, M. Pushkarna, T. Bolukbasi, M. Wattenberg, F. Viégas, and J. Wilson. The what-if tool: Interactive probing of machine learning models. *IEEE transactions on visualization and computer graphics*, 26(1):56–65, 2019.
 - [77] K. Wongsuphasawat, J. A. Guerra Gómez, C. Plaisant, T. D. Wang, M. Taieb-Maimon, and B. Shneiderman. LifeFlow: visualizing an overview of event sequences. In *Proc. CHI*, pp. 1747–1756. ACM, 2011.
 - [78] A. Yamashita, T. Arai, J. Ota, and H. Asama. Motion planning of multiple mobile robots for cooperative manipulation and transportation. *IEEE Transactions on Robotics and Automation*, 19(2):223–237, 2003.